

10/526464  
PCT/JP 03/11298  
24.10.03

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日  
Date of Application: 2002年 9月 5日

REC'D 13 NOV 2003

WIPO

PCT

出願番号  
Application Number: 特願 2002-259737  
[ST. 10/C]: [JP 2002-259737]

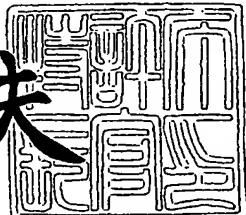
出願人  
Applicant(s): 独立行政法人産業技術総合研究所

PRIORITY DOCUMENT  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

2003年10月 7日

特許庁長官  
Commissioner,  
Japan Patent Office

今井康夫



BEST AVAILABLE COPY

出証番号 出証特 2003-3082844

【書類名】 特許願  
【整理番号】 218-02147  
【あて先】 特許庁長官殿  
【発明者】  
【住所又は居所】 東京都江東区青梅2-41-6 独立行政法人産業技術  
総合研究所臨界副都心センター内  
【氏名】 夏目 徹  
【発明者】  
【住所又は居所】 東京都新宿区改代町2-403  
【氏名】 中山 洋  
【特許出願人】  
【識別番号】 301021533  
【氏名又は名称】 独立行政法人産業技術総合研究所  
【代表者】 吉川 弘之  
【電話番号】 0298-61-3280  
【その他】 国等の委託研究の成果に係る特許出願（平成13年度新  
エネルギー・産業技術総合開発機構「タンパク質機能解  
析」に係わる委託研究、産業活力再生特別措置法第30  
条の適用をうけるもの）  
【提出物件の目録】  
【物件名】 明細書 1  
【物件名】 要約書 1  
【フルーフの要否】 要

【書類名】 明細書

【発明の名称】 生体高分子自動同定方法

【特許請求の範囲】

【請求項1】 試料中の生体高分子の質量を質量分析方法に基づいて測定する質量測定手順と、

前記質量測定手順によって得られる実測質量値を所定のデータベースと照合させることによって候補分子を検索するデータベース検索手順と、

類似順位スコアの高い任意数の候補分子を選び出す候補分子選出手順と、

候補分子を内部標準として用いて実測質量値を校正する質量値校正手順と、

前記手順により得られた候補分子の校正質量値と理論質量値の相対誤差を算出し、該相対誤差の標準偏差を求める手順と、

該標準偏差から前記データベース検索手順の許容誤差を求める手順と、

前記許容誤差に基づき再度前記データベース検索手順と、を行うことを特徴とする生体高分子自動同定方法。

【請求項2】 前記質量値校正手順は、前記候補分子選出手順で選び出された候補分子の実測質量値と理論質量値の相対誤差を算出し、

理論質量値と相対誤差のプロットに対する最小二乗直線を作成して実測質量値の系統誤差を見積もる手順と、

この系統誤差を全実測値から差し引くことにより実測質量値を校正する手順と、からなることを特徴とする請求項1記載の生体高分子自動同定方法。

【請求項3】 コンピュータシステムを利用することにより、請求項1又は2に記載の生体高分子自動同定方法を構成する各手順を実行できるプログラム情報が格納されたことを特徴とする情報記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、質量分析方法を用いた生体高分子同定技術に関する。より詳しくは、質量分析方法によって得られる質量データの精度向上を目的とする生体高分子自動同定方法に関する。

**【0002】****【従来の技術】**

質量分析方法は、試料分子をイオン化した後、これを質量／電荷の比 ( $m/z$ ) に従って分離し検出を行う機器分析法であって、得られた質量スペクトルから定性を、イオン量から定量を行うことができる。

**【0003】**

この分子質量の測定に用いる質量分析計（以下、「MS」（mass spectrometer）と称する。）は、大別すると、試料のイオン化を行うための「イオン化部（イオン源）」と、イオンを質量／電荷の比である  $m/z$  ( $m$ : 質量、 $z$ : 電荷数) に従って分離するための「アナライザ」と、分離されたイオンの「検出部（検出器）」及び「データ解析部」と、から構成されている。

**【0004】**

前記質量分析計を用いて試料分子の質量分析に当たっては、測定開始前に質量分析計の校正（キャリブレーション）を行う必要がある。具体的には、温度変化や電圧の精度、電気回路ノイズ等の原因により、質量分析計の測定に誤差が発生する場合があるので、測定開始前には、クロマトグラフ等を質量分析計から一旦取り外した上で、質量分析計に所定の質量校正用標準物質を導入して実測質量値を得、この実測質量と既知の理論質量値と比較することによって、質量値に系統誤差が発生しないように予め装置を調整する校正作業（外部標準法によるキャリブレーション作業）を行う必要がある。

**【0005】**

更に、高精度の質量値を得るためにには前記外部標準法によるキャリブレーション作業に加え、既知物質を試料に混合して質量測定し、その質量値にもとづいて実測質量値を調整する校正作業（内部標準法によるキャリブレーション作業）を行う必要がある。

**【0006】**

そして、一般に、この質量分析計（タンデム質量分析計を含む。以下同様。）を用いて行うペプチドやタンパク質等の生体高分子同定方法においては、質量分析によって得られる未知試料分子の実測質量値を、10万種程度の分子の一次構

造又は配列が予め格納されているデータベース（ライブラリー）と照合させて検索（サーチ）し、構造から算出した予想リファレンス（標準）スペクトルの中から測定対象の未知試料分子のスペクトルと類似したものを順位（スコア）付けして選びしていく手順、即ちデータベース検索（又はライブラリーサーチ）を行って、候補分子をリストアップして絞り込み、最終的に未知試料分子の同定を行う。

#### 【0007】

##### 【発明が解決しようとする課題】

しかしながら、上記した質量分析計の校正作業（キャリブレーション作業）は、非常に作業が面倒であって、調整時間もかかることから、従来の質量測定作業における作業効率を低下させる主原因であった。即ち、従来は、質量分析計の連続運転（校正作業なしの運転）による効率の良い測定作業を実施することができなかった。また、複数台の質量分析計を用いた測定系においては、各装置について外部標準による校正作業（キャリブレーション作業）を行ったとしても、各装置の精度、信頼性を一元化することは極めて困難であるという問題があった。

#### 【0008】

外部標準キャリブレーションの場合、従来の上記データベース検索の手順では、外部環境の影響によって発生する質量分析計自体の誤測定による影響を測定データから排除することはできなかった。特に測定環境の微妙な温度変化（0.2℃ぐらいの変化）で生じる測定誤差も時として無視できないものとなっていた。

#### 【0009】

また、従来の内部標準キャリブレーションによって複雑な生体高分子混合物を測定する場合は、内部標準物質と試料由来のイオンシグナルが重なってしまい、そのイオンを分析できないため、内部標準として試料に入れる物質の種類や濃度の選択はとても難しかった。高い質量精度を広い質量範囲で実現するためには、何点もの内部標準物質を導入する必要があった。

#### 【0010】

更には、従来は、同定の信頼性が低かったので、その結果を一つ一つ人間が確認しなければならなかった。ところが、近年の質量分析計の発達により、より複雑

な生体高分子混合物の直接分析が可能になってきたため、データが大量化し、一つ一つのデータを人間が目で確認することが不可能になってきたことから、複雑な生体高分子混合物を分析対象とする信頼性の高い自動同定手法の開発が要請されていた。

#### 【0011】

そこで、本発明は、測定開始前の質量分析計の校正作業、あるいはサンプルに予め内部標準を添加することを不要とともに、データ処理のみに基づいた高精度で信頼性の高い生体高分子自動同定方法を提供することを目的とする。

#### 【0012】

##### 【課題を解決するための手段】

上記技術課題を解決するために、本発明では、以下の（1）～（7）の手順を少なくとも備える生体高分子自動同定方法を提供する。

#### 【0013】

（1）試料中の生体高分子の質量を質量分析方法に基づいて測定する質量測定手順。 （2）前記質量測定手順によって得られる実測質量値を所定のデータベースと照合させることによって候補分子を検索するデータベース検索手順。 （3）類似順位スコアの高い任意数の候補分子を選び出す候補分子選出手順。 （4）候補分子を内部標準として用いて実測質量値を校正する質量値校正手順。 （5）前記手順により得られた候補分子の校正質量値と理論質量値の相対誤差を算出し、該相対誤差の標準偏差を求める手順。 （6）該標準偏差から前記データベース検索手順の許容誤差を求める手順。 （7）前記許容誤差に基づき再度前記データベース検索手順。なお、上記「データベース」は、分子構造あるいは配列データベースを意味する。

#### 【0014】

ここで、上記（4）の質量値校正手順は、候補分子選出手順で選び出された候補分子の実測質量値と理論質量値の相対誤差を算出し、理論質量値と相対誤差のプロットに対する最小二乗直線（「 $y = a \times M + b$ 」の式で示される直線。Mは理論質量値。）を作成して実測質量値の系統誤差を見積もる手順と、この系統誤差を全実測値から差し引くことで、実測質量値を校正する手順を採用することが

できる。

### 【0015】

例えば、飛行時間型質量分析計の場合において、候補分子の系統誤差を上記最小二乗直線から求める。この系統誤差を全実測値から差し引く。具体的には、 $(X_c - M) / M = (X - M) / M - (aM + b)$  [Xは実測質量値、X<sub>c</sub>は校正質量値、Mは理論質量値]からなる式を変形し、式： $X_c = X - M(aM + b)$ を得る。

### 【0016】

ここで、理論質量値Mは、候補分子については与えられているが、全ての実測値について与えられている訳ではない。このため、全実測値を校正するためには、上記式のM (aM + b) の項を実測値で近似させる必要がある。a, bの値は、一般にX、X<sub>c</sub>と比較して非常に小さいため、M (aM + b) ≈ X<sub>c</sub> (aX + b) とできる。これを上記式に代入して、 $X_c = X - X_c (aX + b)$ を得る。これを変形し、 $X_c = X / (1 + (aX + b))$  なる式を得て、この式を用いて、全ての実測値を質量校正する。

### 【0017】

上記した本発明に係る生体高分子自動同定方法によれば、複雑な生体高分子混合物を対象として、データ処理のみにより、非常に高精度な質量値を得ることができる。得られる質量値の精度が高いと、より一義的に生体高分子を特定、同定することが可能となる。即ち、本発明は、複雑な生体高分子混合物を分析対象とする信頼性の高い自動同定手法を提供できる。

### 【0018】

次に本願では、コンピュータシステムを利用することにより前記生体高分子自動同定方法を構成する各手順を実行できるプログラム情報が格納されているCD-ROMその他の情報記録媒体を提供する。

### 【0019】

上記した手段によれば、測定開始前の質量分析計の校正作業、あるいはサンプルに予め内部標準を添加することを不要とすることができます。また、データ処理のみに基づいた高精度で信頼性の高い生体高分子自動同定方法を実施することが

できる。

### 【0020】

#### 【発明の実施の形態】

本発明に係る生体高分子自動同定方法の好適な一実施形態について説明する。

なお、本発明は、以下の実施形態に限定されることはない。

### 【0021】

まず、試料中の未知生体高分子の質量を、目的に応じた慣用の質量分析方法に基づいて測定し、実測質量値Xを得る。質量分析方法は、例えば、タンデム質量計を用いることができる。タンデム質量分析計は、アナライザーをタンデムに複数台結合した構成を備える質量分析計であって、詳しくは、最初のアナライザーで混合物中の特定のイオン（親イオン）を選択し、次のアナライザーで選択したイオンと不活性気体との衝突乖離を行い、最後のアナライザーで乖離した内部構造情報を示すイオン（生成イオン）を質量分析する構成を備える。

### 【0022】

前記質量測定手順によって得られた実測質量値Xを、慣用のデータベース検索エンジンが読み込める形式（2値ファイル。質量値と強度。）に変換した上で、その質量値既知の分子が多数記録されたデータベースと照合させて、前記未知生体高分子に該当する可能性のある候補分子の検索を行う。

### 【0023】

なお、上記する実測質量値Xの形式変換は、質量分析計メーカーから一般に提供されている慣用のMasslynx (Micromass社) 等のソフトウェアを適宜用いることによって行うことができ、データベース検索は、市販のMascot (Matrix Science社) 等のデータベースソフトウェアを用いて好適に行うことができる。

### 【0024】

前記データベース検索手順の結果から、類似順位スコアの高い任意数の候補分子（のセット）を選び出す。なお、セットの大きさnは、統計的処理が出来る程度の任意数である。

### 【0025】

続いて、前記した候補分子選出手順によって選び出されてきた各候補分子の実測質量値Xと理論質量値Mの相対誤差Eを、次式（1）に従って算出する。

【0026】

【数1】

$$E = (X - M) / M \dots \dots \dots (1)$$

【0027】

続いて、前記手順によって得られた相対誤差Eの平均値 $m_E$ を次式（2）に基づいて算出する。

【0028】

【数2】

$$m_E = \Sigma (E) / n \dots \dots \dots (2)$$

【0029】

また、前記相対誤差Eの標準偏差 $s_E$ を次式（3）に基づいて算出する。この標準偏差により、候補分子を内部標準として用いることが妥当かどうかを判定する。なお、 $s_E < m_E$ であれば、校正は有効である。

【0030】

【数3】

$$s_E = \{ \Sigma (E - m_E)^2 / (n - 1) \}^{(1/2)} \dots \dots \dots (3)$$

【0031】

次に系統誤差の大きさを見積もり、これを実測質量値Xから差し引くことにより、校正質量値 $X_c$ を得る。例えば、飛行時間型質量分析計の場合において、候補分子の相対系統誤差は、以下の手順で理論質量値と相対誤差のプロットに対する「最小二乗直線 $y = a x + b$ 」から求めることができる。候補分子の校正後の相対誤差 $E_c = (X_c - M) / M$ とすると、 $E_c = E - (a M + b)$ 。したがって、

【0032】

【数4】

$$(X_c - M) / M = (X - M) / M - (a M + b) \dots \dots \dots (4)$$

[Xは実測質量値、 $X_c$ は校正質量値、Mは理論質量値]

## 【0033】

具体的には、上記（4）式を変形して、次式（5）を得る。

## 【0034】

## 【数5】

$$X_c = X - M (aM + b) \dots \dots \dots (5)$$

## 【0035】

ここで、理論質量値は、候補分子については与えられているが、全ての実測値について与えられている訳ではない。このため、全実測値を校正するためには、上記式（5）の「 $M (aM + b)$ 」の項を実測値で近似させる必要がある。 $a$ 、 $b$ の値は、一般に $X$ 、 $X_c$ と比較して非常に小さいので、 $M (aM + b) \approx X_c (aX + b)$ とできるから、これを上記式（6）に代入して、次式（6）を得る。

## 【0036】

## 【数6】

$$X_c = X - X_c (aX + b) \dots \dots \dots (6)$$

## 【0037】

この式（6）を変形式である次式（7）に基づいて、全ての実測値を質量校正する。

## 【0038】

## 【数7】

$$X_c = X / (1 + (aX + b)) \dots \dots \dots (7)$$

## 【0039】

なお、前記最小二乗直線における「 $b$ 」と「 $a$ 」は、それぞれ次式（8）、（9）によって求めることができる。

## 【0040】

## 【数8】

$$b = \sum \{ (M - m_M) \times (E - m_E) \} / \sum \{ (M - m_M)^2 \} \dots (8)$$

## 【0041】

## 【数9】

$$a = m_E - b \times m_M \dots \dots \dots (9)$$

【0042】

さらに、 $m_M$ は、候補分子の理論質量値Mの平均値であって、次式(10)によつて求めることができる。

【0043】

【数10】

$$m_M = \sum (M) / n \dots \dots \dots (10)$$

【0044】

理論質量値Mの相対誤差 $E_c$ は、次式(11)によつて求めることができる。

【0045】

【数11】

$$E_c = E - y \dots \dots \dots (11)$$

【0046】

続いて、候補分子について得られた相対誤差 $E_c = (X_c - M) / M$ の平均値 $m_{E_c}$ 及び標準偏差 $S_{E_c}$ を、それぞれ次式(12)、(13)に基づいて求めること。

【0047】

【数12】

$$m_{E_c} = \sum (E_c) / n \dots \dots \dots (12)$$

【0048】

【数13】

$$S_{E_c} = \{ \sum (E - m_{E_c})^2 / (n - 1) \}^{(1/2)} \dots \dots \dots (13)$$

【0049】

求めた平均値 $m_{E_c}$ から校正を評価する。理想的には $m_{E_c} = 0$ となる。求めた標準偏差 $S_{E_c}$ からデータベース検索に用いる許容誤差 $T_c$ を次式(14)に基づいて算出することによって、一連の校正(キャリブレーション)手順を完了する。

【0050】

【数14】

$T_c = K \times S_E_c \dots \dots \dots \quad (14)$

[ $K = 1.5 \sim 3.0$ ]

#### 【0051】

なお、 $K$ は、質量値の信頼区間を指定するための経験的な定数を示す。この $K$ 値は、データベース検索に用いるソフトウェアの精度に応じて適宜決定できる。データベース検索ソフトウェアの同定性能が高いほど、99.7%の信頼区間である $K=3$ に近づけることが出来る。なお、M a s c o t (Matrix Science社)のデータベースソフトウェアの場合では、経験的に $K=1.5$ を採用できる。

#### 【0052】

得られた前記許容誤差 $T_c$  ( $T_{c1}$ )に基づいて、再度同様のデータベース検索を行う。必要に応じて、上記した一連の校正及びデータベース検索を複数回繰り返すことによって、許容誤差 $T_c$ の範囲を徐々に狭めていき ( $T \rightarrow T_{c1} \rightarrow T_{c2} \rightarrow \dots$ )、候補分子の選出精度を高める。なお、前記 $T_{c1}$ は一回目の校正作業によって得られた許容誤差を示し、 $T_{c2}$ は二回目の校正作業によって得られた許容誤差を示す。

#### 【0053】

これにより、候補分子同定の確度を高めていくことができる。即ち、未知試料分子の同定精度を向上させることができる。

#### 【0054】

以上説明した手順を所望のコンピュータプログラム情報に加工し、このプログラム情報をCD-ROM、フロッピー（登録商標）ディスクなどの各種情報記録媒体、コンピュータハードウエア、サーバ等に格納し、所望のコンピュータシステムやコンピュータネットワーク（情報通信技術）を介して、該プログラムを実行可能に工夫することができる。

#### 【0055】

##### 【発明の効果】

本発明によれば、測定開始前の質量分析計の校正作業、あるいはサンプルに予め内部標準を添加することを不要とすることができる。そのため、質量分析計の連続運転（校正作業による中断のない運転）が可能となる。その結果、作業者は煩わし

い装置調整作業から開放され、分子同定作業の効率を向上させることができる。

### 【0056】

また、質量分析計自体の誤差の影響を排除し、データ処理のみに基づいた高精度で信頼性の高い生体高分子自動同定方法を実施することができ、そして、複数台の質量分析計を用いた測定系では、各質量分析計から得られるデータ精度の一元化を達成できるので、未知試料分子の誤同定を確実に防止することができる。

【書類名】要約書

【要約】

【課題】 (1) 測定開始前の質量分析計の校正作業、あるいはサンプルに予め内部標準を添加することを不要とすること。 (2) データ処理のみに基づいた高精度で信頼性の高い生体高分子自動同定を可能とすること。

【解決手段】 試料中の生体高分子の質量を質量分析方法に基づいて測定する質量測定手順と、前記質量測定手順によって得られる実測質量値Xを所定のデータベースと照合させることによって候補分子を検索するデータベース検索手順と、類似順位スコアの高い任意数の候補分子を選び出す手順と、前記候補分子を内部標準として用いて実測質量値Xを校正する質量値校正手順と、前記手順により得られた候補分子の校正質量値X<sub>c</sub>と理論質量値Mの相対誤差Eを算出し、該相対誤差Eの標準偏差s<sub>E</sub>を求める手順と、この標準偏差s<sub>E</sub>から前記データベース検索手順の許容誤差T<sub>c</sub>を求める手順と、この許容誤差T<sub>c</sub>に基づいて再度前記データベース検索を行う手順と、を備える生体高分子自動同定方法を提供する。

## 認定・付加情報

特許出願の番号 特願2002-259737  
受付番号 50201327181  
書類名 特許願  
担当官 井筒 セイ子 1354  
作成日 平成14年 9月 9日

## &lt;認定情報・付加情報&gt;

【提出日】 平成14年 9月 5日

次頁無

特願2002-259737

出願人履歴情報

識別番号 [301021533]

1. 変更年月日 2001年 4月 2日  
[変更理由] 新規登録  
住 所 東京都千代田区霞が関1-3-1  
氏 名 独立行政法人産業技術総合研究所